1996 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

ADAPTIVE NOISE SUPPRESSION USING DIGITAL SIGNAL PROCESSING

Dr. David Kozel, Associate Professor
Engineering Department
Purdue University Calumet
Hammond, Indiana

KSC Colleague - Richard Nelson
Communications/Rf and Audio

Contract Number NASA-NGT10-52605

August 9, 1996

Abstract

A signal to noise ratio dependent adaptive spectral subtraction algorithm is developed to eliminate noise from noise corrupted speech signals. The algorithm determines the signal to noise ratio and adjust the spectral subtraction proportion appropriately. After spectral subtraction low amplitude signals are squelched. A single microphone is used to obtain both the noise corrupted speech and the average noise estimate. This is done by determining if the frame of data being sampled is a voiced or unvoiced frame. During unvoiced frames an estimate of the noise is obtained. A running average of the noise is used to approximate the expected value of the noise. Applications include the Emergency Egress Vehicle and the Crawler-Transporter.

1. Introduction

It is desired to incorporate adaptive noise suppression into the communications equipment on the Emergency Egress Vehicle and the Crawler-Transporter. In the case of the Emergency Egress Vehicle, people are fixed relative to the noise source. The spectral content of the noise source changes as a function of the speed of the vehicle and its engine. In the case of the Crawler-Transporter the people can move relative to the Crawler-Transporter. Thus, the noise a person hears will vary with their location relative to the Crawler-Transporter and if the hydrolic leveling device on the Crawler-Transporter is being used. Due to the varying nature of the noise, an adaptive algorithm is necessary for both applications. Furthermore, the noise frequencies produced by both applications are in the voice band range. Thus, standard filtering techniques will not work. A signal to noise ratio dependent adaptive spectral subtraction algorithm is developed to eliminate the noise. OIS-D microphones are used. OIS-D microphones have noise suppression of a mechanical nature, which provides approximately 15dB of noise suppression. This is sufficient to provide a signal to noise ratio favorable enough for spectral subtraction to perform very well.

2. Spectral Subtraction

The additive noise model used for spectral subtraction assumes that noise corrupted speech is composed of speech plus additive noise.

$$x(t)=s(t) + n(t) \tag{1}$$

where:

- x(t) noise corrupted speech
- s(t) speech
- n(t) noise

Taking the Fourier Transform of equation (1)

$$X(f) = S(f) + N(f) \tag{2}$$

$X(f)$, $S(f)$, and $N(f)$ are complex so they can be represented in polar form

$$|X(f)|e^{j\theta x} = |S(f)|e^{j\theta s} + |N(f)|e^{j\theta n} \tag{3}$$

Solving for the speech

$$|S(f)|e^{j\theta s} = |X(f)|e^{j\theta x} - |N(f)|e^{j\theta n} \tag{4}$$

Since the phase of the noise is in general unavailable, the phase of the noise corrupted speech is commonly used to approximate the phase of the speech. This is equivalent to assuming the noise corrupted speech and the noise are in phase. As a result the speech magnitude is approximated from the difference of the noise corrupted speech and noise magnitudes.

$$\hat{S}(f) = \left|\hat{S}(f)\right| e^{j\theta x} = \left(|X(f)| - |N(f)|\right) e^{j\theta x} \tag{5}$$

The inverse Fourier Transform yields the estimate of the speech.

$$\hat{s}(t) = \mathcal{F}^{-1}\left\{\hat{S}(f)\right\} \tag{6}$$

There are different types of spectral subtraction. The type described above is termed magnitude spectral subtraction, because the magnitude of the noise spectrum at each frequency is subtracted. A derivation for power spectral subtraction is given in [1]. In its most general form spectral subtraction is written as [2]

$$\hat{S}(f) = \left\{|X(f)|^b - \alpha(SNR(f))E\left[|N(f)|^b\right]\right\}^{\frac{1}{b}} e^{j\theta x} \tag{7}$$

The exponent, b, equals 1 for magnitude spectral subtraction and 2 for power spectral subtraction. The proportion of noise subtracted, α, can be variable and signal to noise ratio dependent. In general α is greater than 1 which is termed over subtraction. Over subtraction is used to reduce the distortion caused from approximating the phase. In equation (7), E[*] represents the expected value of [*].

2.1 Limitations of Spectral Subtraction

When using any algorithm, it is important to understand its limitations and restrictions. The phase approximation used in the speech estimate produces both magnitude and phase distortion in each frequency component of the speech estimate. This can be seen in Figure 1 by the vector representation of equation (4) and equation (5) respectively for any one frequency. If the magnitude of the noise, $|N|$, is "small" relative to the magnitude of the corrupted speech, $|X|$, the distortion caused by using the noise corrupted speech phase, θ_x, in place of the noise phase is minimal and unnoticeable to the human ear. Likewise, if the phase of the noise, θ_n, is "close" to the phase of the corrupted speech, θ_x, the resulting error produced by the approximation is

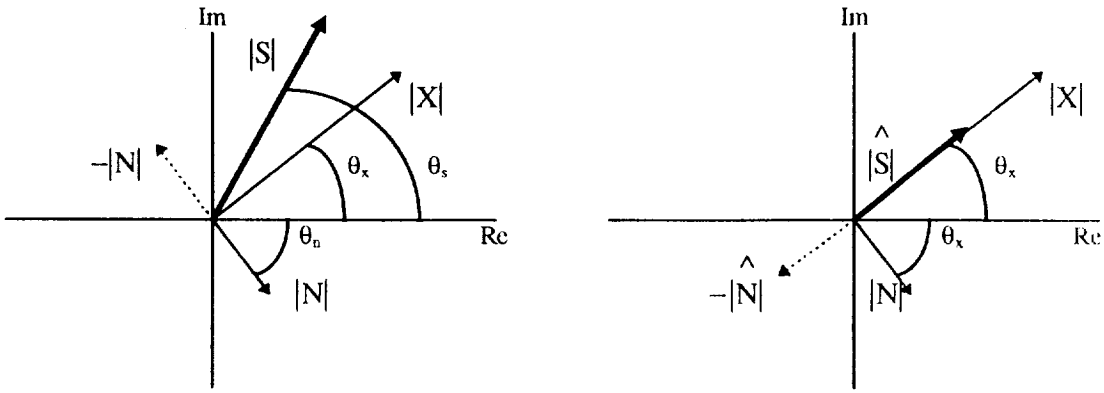

Figure 1. Vector Representation of Equations (4) and (5) Respectively

minimal and unnoticeable to the human ear. Since the relative phase between θ_x and θ_n is unknown and varies with time and frequency, the ratio between the magnitude of the noise corrupted speech and the noise is used as an indication of accuracy.

3. Signal to Noise Ratio Dependent Adaptive Spectral Subtraction Algorithm

A Diagram of the Signal to Noise Ratio Dependent Adaptive Spectral Subtraction algorithm (SNRDASS) is shown in Figure 2. Details of the algorithm are described in the following subsections.

3.1 Framing, Windowing, Zero Padding and Recombining the Signal

The process of windowing, zero padding, and recombination of the signal is shown in Figure 3. The sampled signal is segmented into frames each containing m points. Each frame is multiplied by a triangular window containing m points. This is required since the algorithm uses a Fast Fourier Transform (FFT) which assumes that the signal is periodic relative to the frames. If a window is not used spurious frequencies are produced due to signal levels at the ends of each frame not being equal. As a result of windowing each frame is required to overlap the previous frame in time by 50 percent. This allows the two triangular windowed components to add to the original signal when recombined. If a window type other than a triangular window is used the addition of frames can produce oscillation errors of up to approximately 9 percent of the original amplitude in the recombined signal. Spectral subtraction can be considered as a time varying filter[3] which can vary from frame to frame.

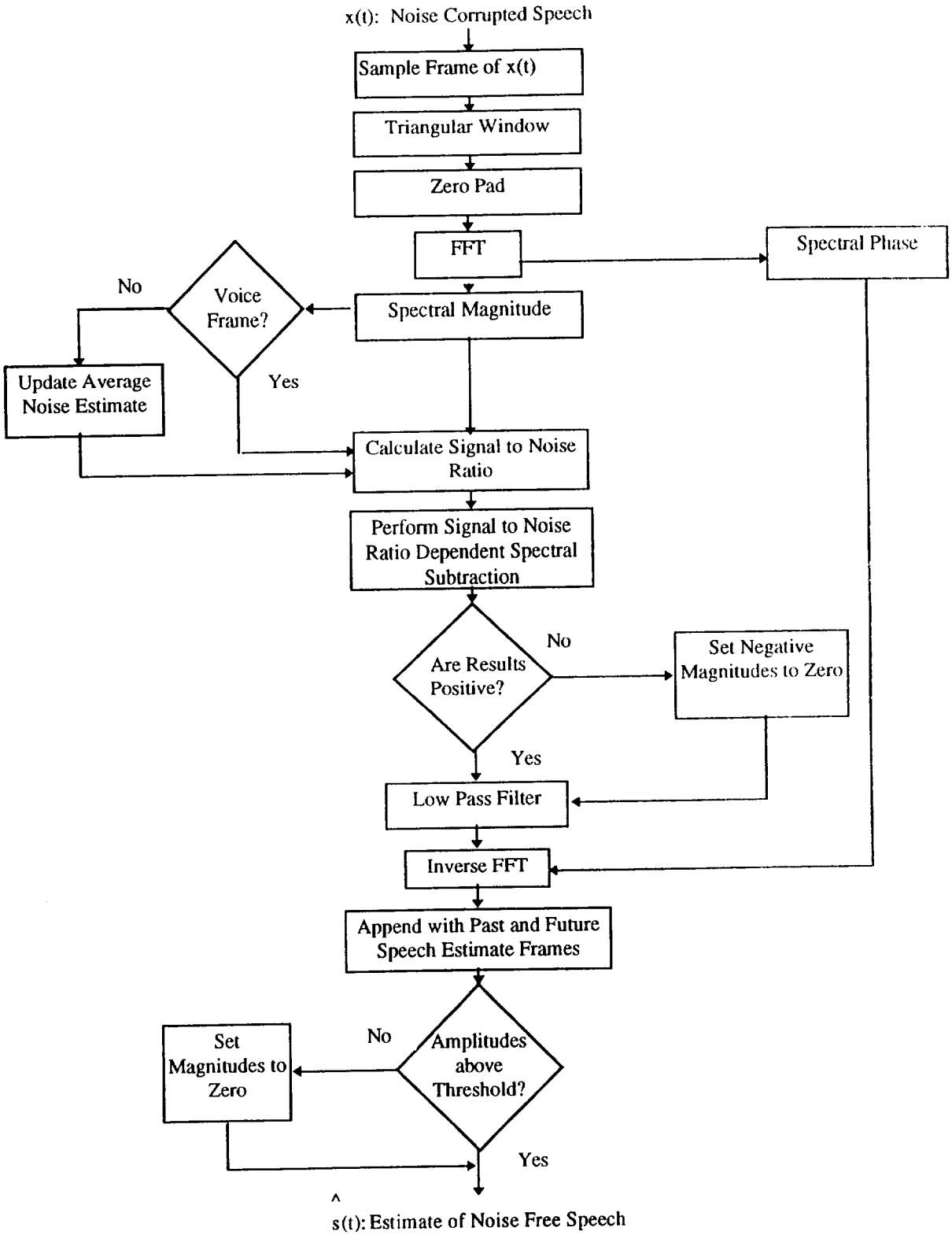

Figure 2 Signal to Noise Ratio Dependent Adaptive Spectral Subtraction Algorithm

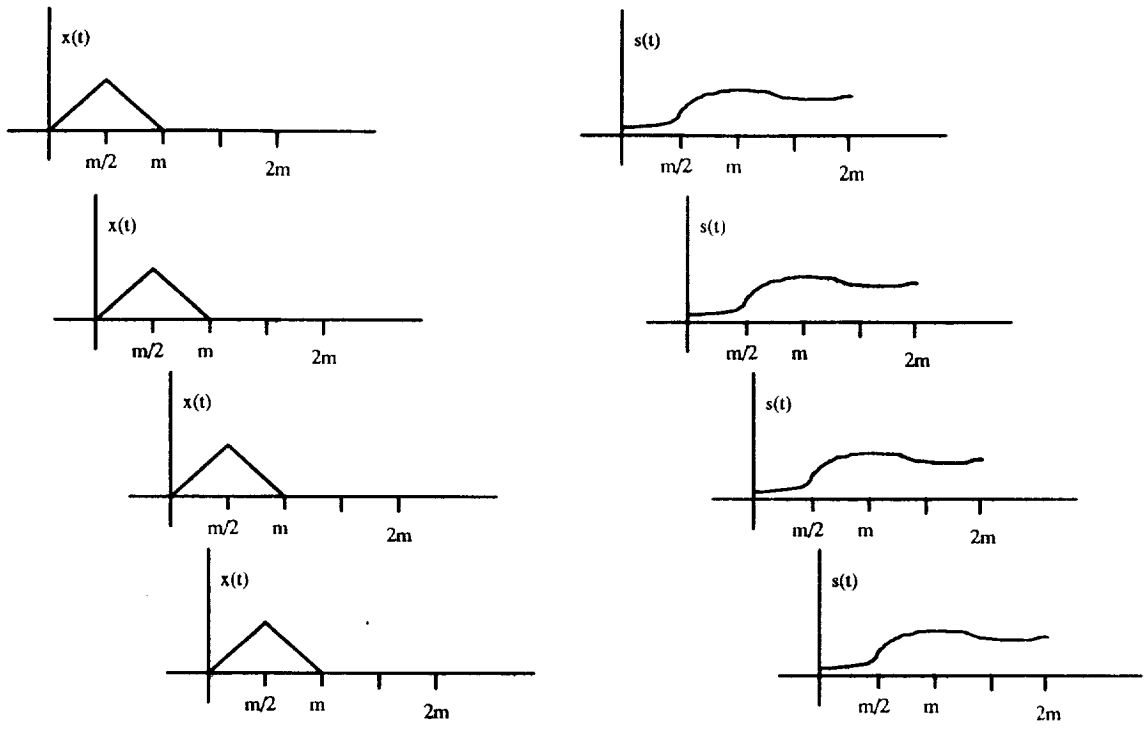

Figure 3. Windowing, Zero Padding, and Signal Recombination

$$\hat{S}(f) = \left| \hat{S}(f) \right| e^{j\theta x} = |H(f)||X(f)|e^{j\theta x} = \left|1 - \frac{|N(f)|}{|X(f)|}\right| |X(f)| e^{j\theta x} = \left(|X(f)| - |N(f)|\right) e^{j\theta x} \tag{8}$$

Since the filter is obtained from the corrupted speech and noise it has a length of m points. The length of the time domain response of such a filter will be 2m-1. To eliminate the effects of circular convolution the windowed signal must be zero padded by m points to a total length of 2m points [4].

Since there is a 50 percent overlap in each frame only m/2 points of new information is obtained. But since the response last for 2m points four output frames overlap in time and must be combined to produce the correct output for each frame. This is a variation of the overlap add method described in [5].

3.2 Voice Frame Recognition

Once the signal has been windowed and zero padded the FFT is taken. The resulting magnitude and phase of the signal spectrum are determined. The phase is set aside for recombination with the spectral subtracted magnitude. The magnitude of the signal spectrum is used to determine if

the frame contains voice or is voice free. This is done by comparing the maximum value of the signal magnitude spectrum with a proportion, γ, of the maximum value of the average noise magnitude spectrum.

$$\text{if } \max(|X(kf)|) > \gamma \max(\overline{|N(kf)|}) \quad \text{for } k = 1,\ldots,m \quad \text{then consider frame to be voiced} \tag{9}$$

The proportion, γ, can be initialized by comparing the maximum magnitude of a known voice frame to the maximum magnitude of the average noise.

The average magnitude spectrum for the noise is obtained by the following procedure: When the algorithm is first being initialized an initial noise only sequence of frames must be obtained to get a baseline on the average magnitude spectrum of the noise.
For frame 1 of the initial noise only sequence:

$$\overline{|N(kf)|} = |X(kf)| \quad \text{for } k = 1,\ldots, m \tag{10}$$

for other frames of the initial noise only sequence:

$$\overline{|N(kf)|} = \delta \overline{|N(kf)|} + (1 - \delta)|X(kf)| \quad \text{for } k = 1,\ldots, m \tag{11}$$

where $0.70 \le \delta \le 0.95$.
Once the initial average noise estimate is obtained, from a known noise only test sequence, each frame of signal is checked for voice using equation (9). If equation (9) is not satisfied the frame is considered unvoiced and equation (11) is used with a predetermined value for δ that is in the specified range. In general δ determines how quickly the noise estimate can vary. The technique is simple, but works well, since voice frames are generally strong in specific frequencies due to excitation of the vocal cords. A more computationally intensive algorithm is used in [6].

3.3 Adaptive Signal to Noise Ratio Dependent Spectral Subtraction Algorithm

After the average noise magnitude spectrum is updated according to the rules outlined in Section 3.2, the magnitude spectrum of the signal and the average noise magnitude spectrum are used to perform spectral subtraction. The signal to noise ratio dependent proportion, α, is determined using the following equation:

$$\alpha = \frac{\eta \sum_{k=1}^{m} \overline{|N(kf)|}}{\sum_{k=1}^{m} |X(kf)|} \tag{12}$$

When the algorithm is first initialized η is determined by testing a signal frame that is known to contain voice. η is chosen such that α is approximately 1.78 in the voiced frames. Once α is determined spectral subtraction is performed using

$$\left| \hat{S}(kf) \right| = | X(kf) | - \alpha \overline{| N(kf) |} \qquad \text{for } k = 1,\ldots,2m \tag{13}$$

If any of the estimates for $\left| \hat{S}(kf) \right|$ are negative, they are set to zero. $\left| \hat{S}(kf) \right|$ is then lowpass filtered eliminate musical noise which is generally high frequency. The lower the 3dB frequency of the filter the more noise and speech eliminated. Depending on the results achieved from spectral subtraction this step may not provide appreciable improvement. Furthermore, the smoothing filter discussed in Section 4 will achieve some of the same results. If calculation time becomes an issue for real time implementation the lowpass filter can be omitted from the algorithm.. After lowpass filtering, the phase of the noise corrupted speech, θ_x, is combined with the magnitude of the estimate of the speech and the inverse FFT is taken. This provides one of the four offset output frames that must be combined using the overlap add method described in Section 3.1. The summing provides an averaging effect for phase errors which reduces the error. But, since each frame is filtered by a different transfer function, see equation (8), summing frames also produces discontinuities in the response causing musical noise.

3.4 Low Level Signal Squelching

The low level signal squelching algorithm looks at three frames of estimated speech: the past, present, and future frames. Future frame estimates of speech are obtained by delaying the speech estimate for one frame before being output. Thus, the signal to noise ratio dependent spectral subtraction algorithm is actually calculating the future output, while the present output is being held in a buffer to determine if low level squelching is required, and the past frame is being output through the D/A. The algorithm is described by the following equation:

$$\begin{aligned} &\text{if} \quad \left|\hat{S}(kf,i)\right| < \mu \ \max(\overline{|N(kf,L)|}) \qquad \text{for } k = 1,\ldots,m/2, \quad \text{and } i = L-1,\ L,\ L+1 \\ &\text{then} \ \left|\hat{S}(kf,L)\right| = 0 \quad \text{for } k = 1,\ldots,m/2 \end{aligned} \tag{14}$$

where μ is a user discretion proportion. A frequency domain squelching technique is given in [7].

4. Overview

A block diagram of the adaptive noise suppression system is shown in Figure 4. Noise or noise corrupted speech enters the microphone. A high gain amplifier is used to bring the voiced signal up to the ±2.5 Volt range used by the Analog to Digital (A/D) converter. Before entering the A/D converter the signal passes through an anti aliasing lowpass filter with 3dB attenuation at 3KHz and 30dB attenuation at 5.9KHz. It is then sampled by the A/D converter using 12 bit

resolution and a 14.925KHz sampling rate. At this point the Digital Signal Processor (DSP) performs noise suppression using signal to noise ratio dependent adaptive spectral subtraction. Next, the digital signal is converted back to an analog signal at a rate of 14.925KHz using the Digital to Analog converter. It is then sent through a smoothing filter, which for the data obtained in the testing was a lowpass Bessel filter with a 3dB frequency of 3KHz. This can be replaced with a voice band filter, which is a bandpass filter with low and high 3dB passband frequencies of 300 and 3KHz respectively. If the voice band filter does not have good damping characteristics the smoothing filter is necessary or transients will be produced from the step discontinuities resulting from the D/A conversion. After the voice band filter the signal is modulated and transmitted by the communication device.

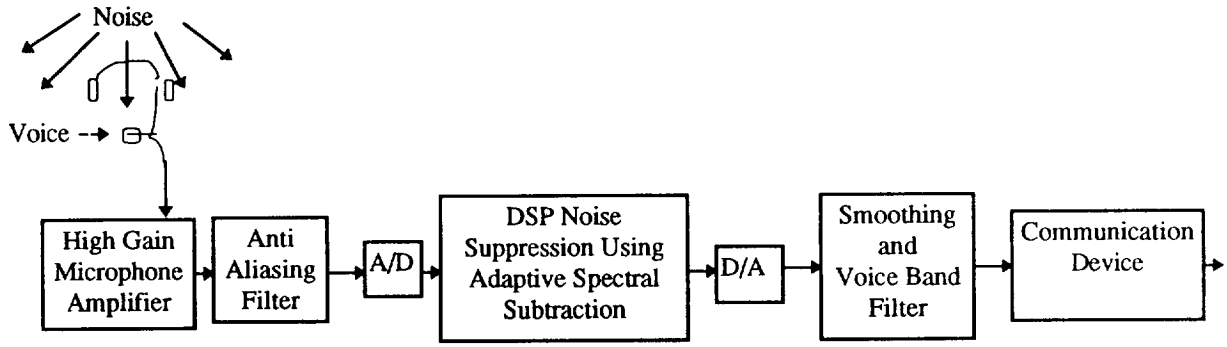

Figure 4. Adaptive Noise Suppression Block Diagram

5. Application

The emergency egress vehicle is shown in Figure 5. Basically it is an M113 tank, which is used to

Figure 5. Exterior and Interior of Emergency Egress Vehicle

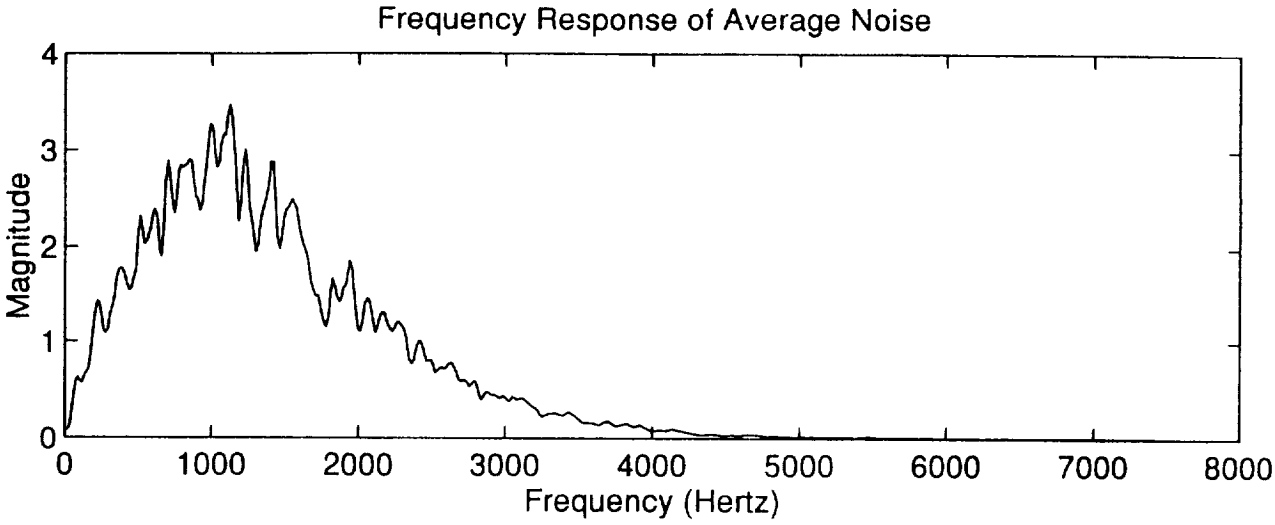

Figure 6. Frequency Response of the Average Noise

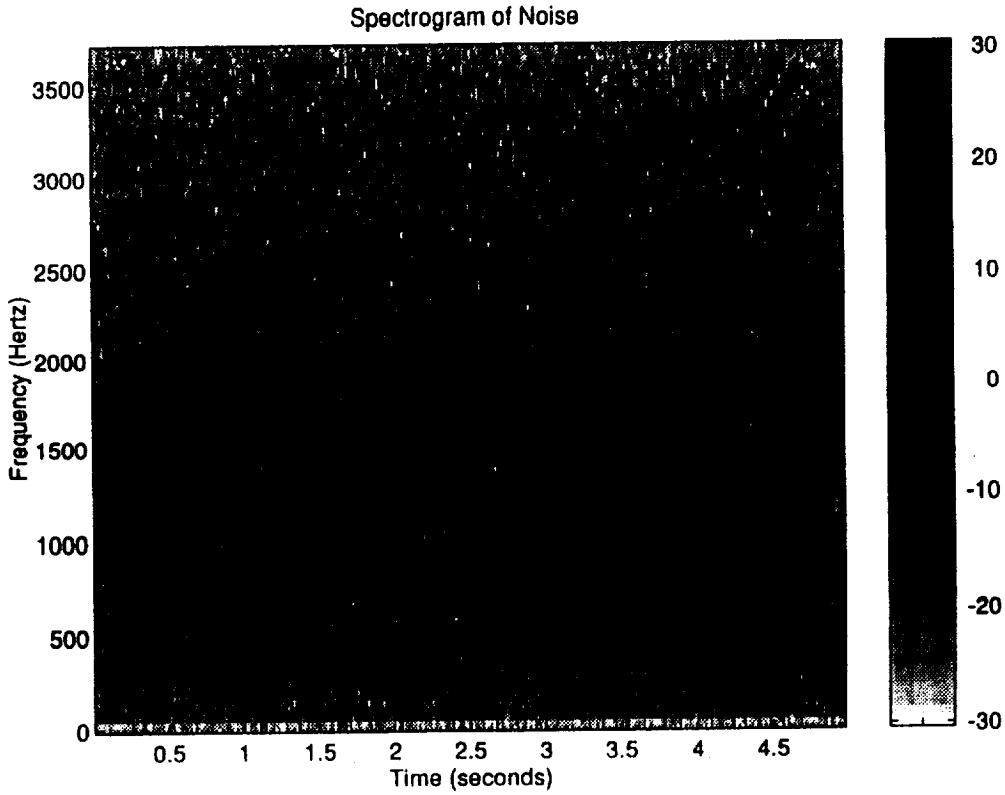

Figure 7. Spectrogram of the Noise

evacuate the astronauts if an emergency situation arises during launch. The noise level inside the vehicle is 90 decibels with the engine running and 120-125 decibels once the vehicle starts moving. As a result, it is impossible to hear what the M113 crew is saying during a rescue operation. The headsets used by the rescue crew have OIS-D microphones which have noise suppression of a mechanical nature, which provides 15 decibels of noise suppression. Furthermore, the frequency response of the microphone attenuates frequencies outside of the voice band range of 300Hz to 3kHz. The frequency response and spectrogram of the noise as obtained through the microphone are shown in Figure 6 and Figure 7 respectively. The decibel scale to the right in Figure 7 is determined using

$$|N_{dB}(f)| = 20\log_{10}|N(f)| \tag{15}$$

Where $|N(f)|$ is the FFT of the noise n(t) obtained from the A/D converter. Thus the spectrogram levels are relative to the voltage of the A/D, not absolute voice levels.

From Figure 6 it is apparent that the noise which is input by the microphone is directly in the range of voice band frequencies. Thus using standard filtering techniques to attenuate it will also attenuate speech by the same factor. The spectrogram of the noise shown in Figure 7 shows that the noise is not constant. As each track of the M113 tank hits the ground, the reaction force causes an impulse on the tank which excites its resonant frequencies.

6 Test Results

The signal to noise ratio dependent adaptive spectral subtraction algorithm was tested on the emergency egress vehicle shown in Figure 5 using the following parameter settings:

$m = 512$

$\gamma = 2.0$

$\delta = 0.90$

$\eta = 4.0$

$\mu = 0.025$

The words "test, one, two, three, four, five" were spoken into the microphone. The original sampled signal, signal after spectral subtraction, and signal after squelching are displayed in Figure 8. Spectrograms for the same three conditions are shown in Figure 9. It can be seen from Figures 8 and 9 that a signal to noise ratio of approximately 15dB exist for the original sampled signal. As mentioned in Section 5 the microphones provided approximately 15dB of noise attenuation. This provided a favorable signal to noise ratio, which is required for spectral subtraction to work well. Lowering the gain and talking louder also improved the signal to noise ratio without saturating the voltage limits of the A/D converter. It can be seen from Figures 8 and 9 that spectral subtraction provided approximately 20dB of improvement in the signal to noise ratio. Listening test verified that the noise was virtually eliminated, with little or no distortion due to musical noise.

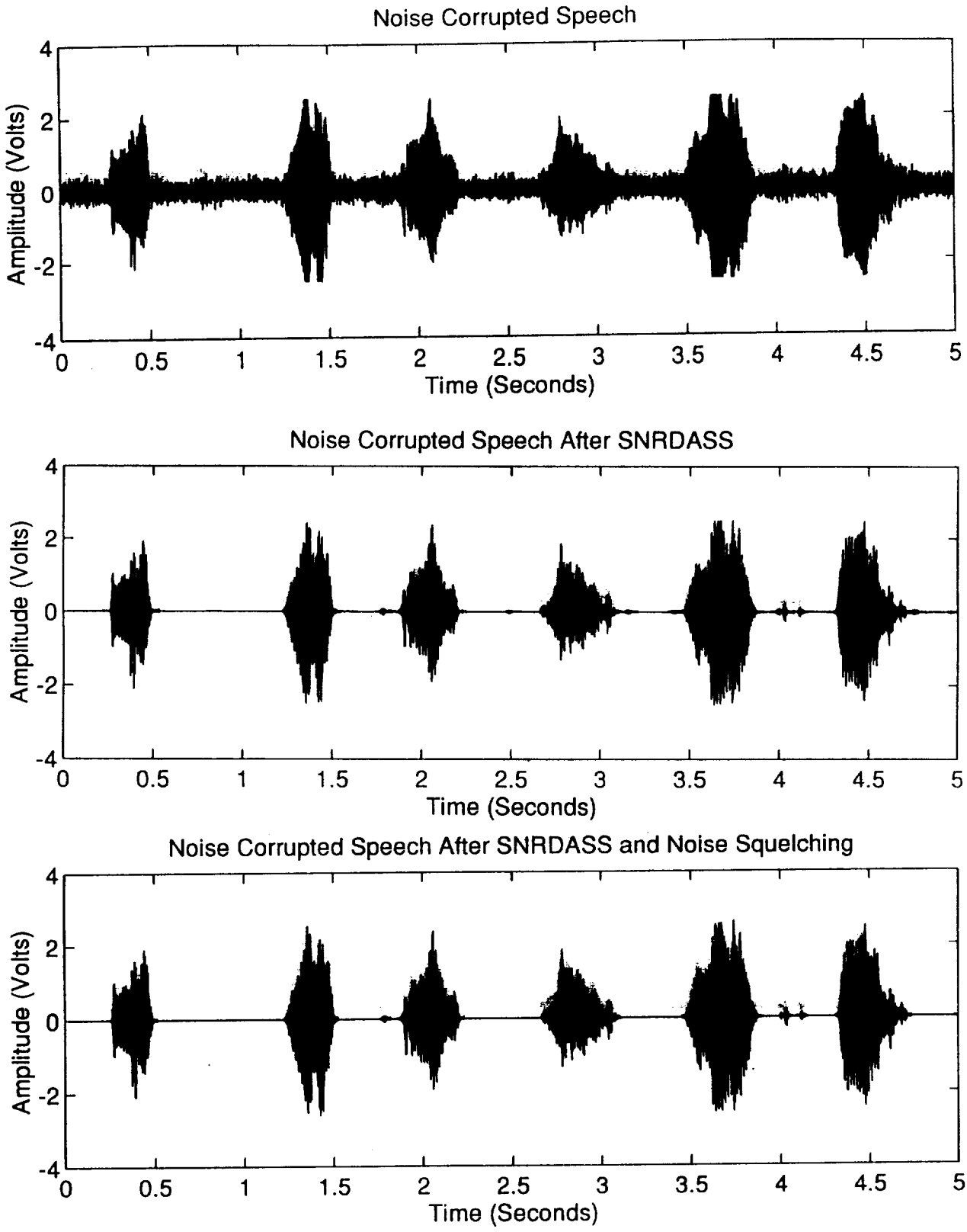

Figure 8. Time Domain Results

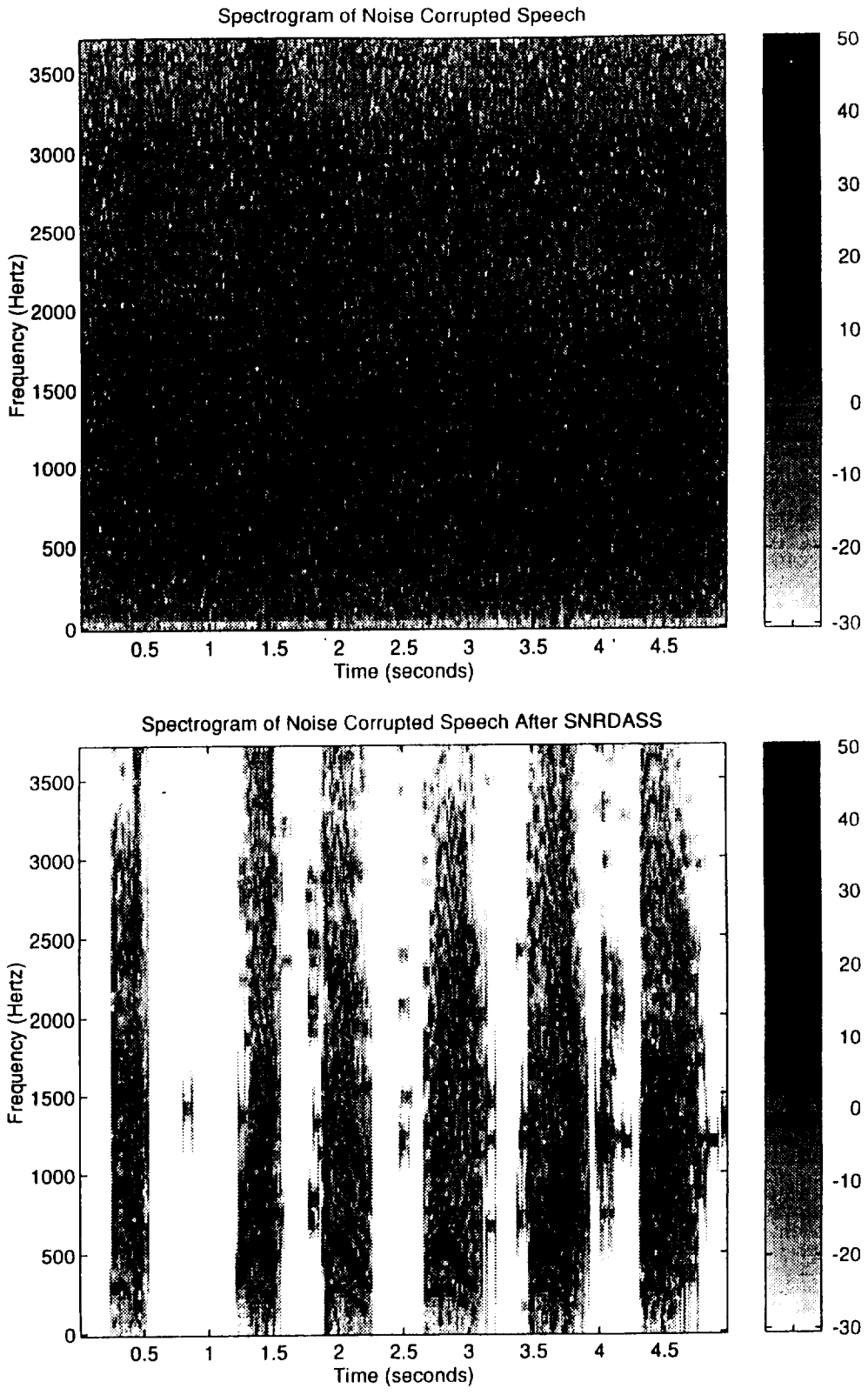

Figure 9. Spectrograms of the Results

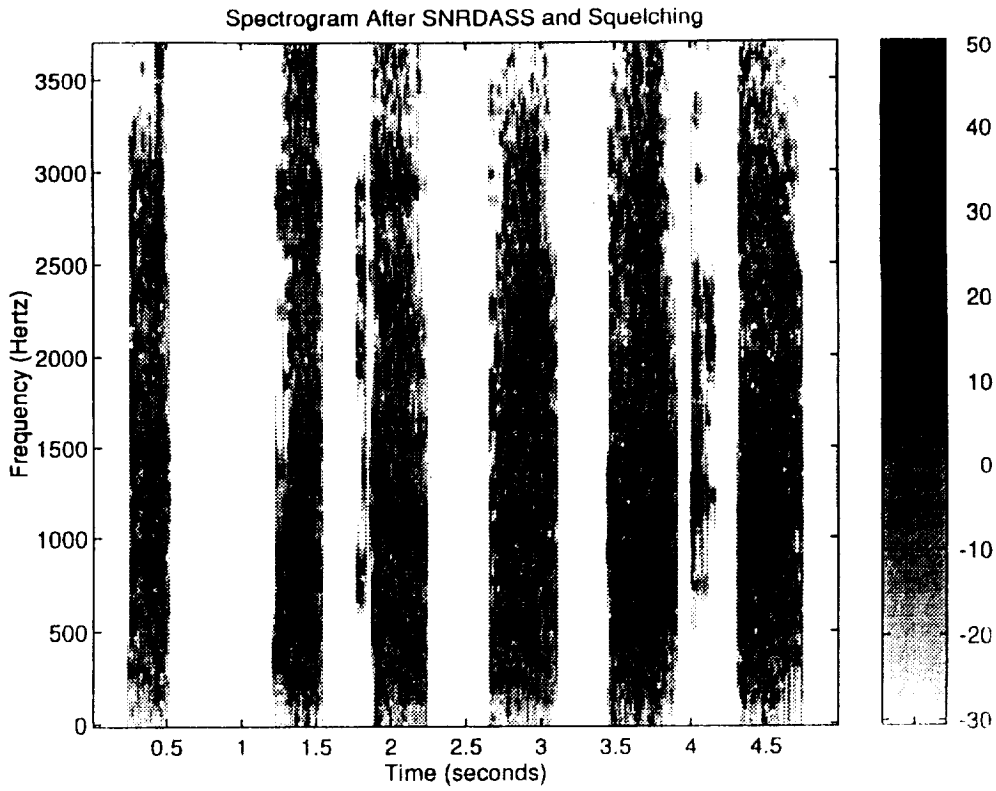

Figure 9 Continued. Spectrograms of the Results

References

[1] McOlash, Scott M., Niederjohn, Russell J., and Heinen, James A., A Spectral Subtraction Method for the Enhancement of Speech Corrupted by Non-White, Non-Stationary Noise, Proc. of the 1995 IEEE 21st International Conference on Industrial Electronics, Control, and Instrumentation, v2, pt2, Nov 6-10, pg 872-877.

[2] Milner, B. P., and Vaseghi, S. V., Comparison of Some Noise-Compensation Methods for Speech Recognition in Adverse Environments, IEE Proceedings: Vision, Image, and Signal Processing, v141, Oct. 5,1994, p280-288.

[3] Arslan, Levent, McCree, Alan, and Viswanathan, Vishu, New Methods for Adaptive Noise Suppression, Proc. of the 1995 20th International Conference on Acoustics, Speech, and Signal Processing, v1, pt1, May 12, 1995, pg 812-815.

[4] Cunningham, Edward P., Digital Filtering: An Introduction,Houghton Mifflin Co., Boston, MA, 1992.

[5] Proakis, John G., and Manolakis, Dimitris G., Digital Signal Processing Principles, Algorithms, and Applications, Second Edition, Macmillan Publishing Co., New York, NY, 1992.

[6] Hirsch, H. G., and Ehrlicher, C., Noise Estimation for Robust Speech Recognition, Proc. of the 1995 20th International Conference on Acoustics, Speech, and Signal Processing, v1, pt1, May 12, 1995, pg 153-156.

[7] Ching, Wee-Soon, and Toh, Peng-Seng, Enhancement of Speech Signal Corrupted By High Acoustic Noise, Proc. of the 1993 IEEE Region 10 Conf. on Computers Communications, Control, & Power Engineering, pt2, Oct 19-21, 1993, pg 1114-1117.